CLAIMS:

1. A method of determining cluster attractors for a plurality of documents, each document comprising at least one term, the method comprising: calculating, in respect of each term, a probability distribution indicative of the frequency of occurrence of the, or each, other term that co-occurs with said term in at least one of said documents; calculating, in respect of each term, the entropy of the respective probability distribution; selecting at least one of said probability distributions as a cluster attractor depending on the respective entropy value.

2. A method as claimed in Claim 1, wherein each probability distribution comprises, in respect of each co-occurring term, an indicator that is indicative of the total number of instances of the respective co-occurring term in all of the documents in which the respective co-occurring term co-occurs with the term in respect of which the probability distribution is calculated.

3. A method as claimed in Claim 1 or 2, wherein each probability distribution comprises, in respect of each co-occurring term, an indicator comprising a conditional probability of the occurrence of the respective co-occurring term in a document given the appearance in said document of the term in respect of which the probability distribution is calculated.

4. A method as claimed in any one of Claims 1 to 3, wherein each indicator is normalized with respect to the total number of terms in the, or each, document in which the term in respect of which the probability distribution is calculated appears.

5. A method as claimed in Claim 1, comprising assigning each term to one of a plurality of subsets of terms depending on the frequency of occurrence of the term; and selecting, as a cluster attractor, the respective probability distribution of one or more terms from each subset of terms.

6. A method as claimed in Claim 5, wherein each term is assigned to a subset depending on the number documents of the corpus in which the respective term appears.

7. A method as claimed in Claim 5 or 6, wherein an entropy threshold is assigned to each subset, the method comprising selecting, as a cluster attractor, the respective probability distribution of one or more terms from each subset having an entropy that satisfies the respective entropy threshold.

8. A method as claimed in Claim 7, comprising selecting, as a cluster attractor, the respective probability distribution of one or more terms from each subset having an entropy that is less than or equal to the respective entropy threshold.

9. A method as claimed in any one of Claims 5 to 8, wherein each subset is associated with a frequency range and wherein the frequency ranges for respective subsets are disjoint.

10. A method as claimed in any one of Claims 5 to 9, wherein each subset is associated with a frequency range, the size of each successive frequency range being equal to a constant multiplied by the size of the preceding frequency range in order of increasing frequency.

11. A method as claimed in any one of Claims 7 to 10, wherein the respective entropy threshold increases for successive subsets in order of increasing frequency.

12. A method as claimed in Claim 11, wherein the respective entropy threshold for successive subsets increases linearly.

13. A computer program product comprising computer program code for causing a computer to perform the method of Claim 1.

14. An apparatus for determining cluster attractors for a plurality of documents, each document comprising at least one term, the apparatus comprising: means for calculating, in respect of each term, a probability distribution indicative of the frequency of occurrence of the, or each, other term that co-occurs with said term in at least one of said documents; means for calculating, in respect of each term, the entropy of the respective probability distribution; and means for selecting at least one of said probability distributions as a cluster attractor depending on the respective entropy value.

15. A method of clustering a plurality of documents, each document comprising at least one term, the method comprising determining cluster attractors in accordance with Claim 1.

16. A method as claimed in Claim 15, comprising: calculating, in respect of each document, a probability distribution indicative of the frequency of occurrence of each term in the document; comparing the respective probability distribution of each document with each probability distribution selected as a cluster attractor; and assigning each document to at least one cluster depending on the similarity between the compared probability distributions.

17. A method as claimed in Claim 16, comprising organising the documents within each cluster by: assigning a respective weight to each document, the value of the weight depending on the similarity between the probability distribution of the document and the probability distribution of the cluster attractor; comparing the respective probability distribution of each document in the cluster with the probability distribution of each other document in the cluster; assigning a respective weight to each pair of compared documents, the value of the weight depending on the similarity between the compared respective probability distributions of each document of the pair; calculating a minimum spanning tree for the cluster based on the respective calculated weights.

18. A computer program product comprising computer program code for causing a computer to perform the method of Claim 15.